Computer-based analysis

Lawrence, John; Visser, Jacky

Published in:
The Routledge Handbook of Language and Persuasion

DOI:
10.4324/9780367823658-37

Publication date:
2022

Document Version
Peer reviewed version

Lawrence, J & Visser, J 2022, Computer-based analysis: Argument mining. in J Fahnestock & RA Harris (eds), The Routledge Handbook of Language and Persuasion. 1 edn, Routledge, London, pp. 525-540. https://doi.org/10.4324/9780367823658-37

Computer-based Analysis: Argument Mining

John Lawrence & Jacky Visser

1 Introduction

The continuously increasing volume of textual data ripe for analysis is driving the development of computational efforts to unlock the wealth of information contained within. Automated techniques for analysing persuasive texts, such as Opinion Mining and Sentiment Analysis, make it possible to identify the attitudes expressed in a piece of text – for example, whether a product review is positive or negative (Pang & Lee, 2008). While these well-established techniques can be effectively used to determine the stance of persuasive texts, they stop short of reconstructing the persuasive means advanced in support of that stance. The automated reconstruction of the argumentative structure of persuasive texts, *Argument Mining*, meets this challenge by reconstructing not only *which* claims are being advanced, but also *why* those particular standpoints are held.

The earliest approaches to argument mining (Palau & Moens, 2009) were aimed at detecting the argumentative parts of a text by first splitting it into sentences and using features of these sentences to classify each as either *Argument* or *Non-Argument*, and then classifying each *Argument* sentence as either *premise* or *conclusion*. Whilst much recent work in this area builds on these concepts and techniques, the range of tasks and technologies available has grown dramatically. The resulting computational tasks can be broadly categorised as:

- identifying argument components, including boundary detection and argument/non-argument classification
- identifying clausal properties, both intrinsic, such as whether the clause is factual or evaluative, and contextual such as whether the clause is the conclusion to an argument
- identifying relational properties, from simple premise/conclusion relationships, to whether a set of clauses forms an instance of an argument scheme

Argument analysis aims to turn unstructured text into structured argument data, giving an understanding, not just of the individual points being made, but of the relationships between them and how they work together to support (or counter) the overall message (van Eemeren et al., 2014). Whilst there is evidence that argument analysis aids comprehension of large volumes of data, the manual extraction of argument structure is a skilled and time consuming process. As a case in point, Robert Horn talking about the argument maps he produced on the debate as to whether computers can think, quotes a student as saying "These maps would have saved me 500 hours of time my first year in graduate school"[1], however Metzinger, 1999 notes that over 7,000 hours of work was required in order for Horn and his team to create those maps.

Although attempts have been made to increase the speed of manual argument analysis, it is impossible to keep up with the rate of textual data being generated across even a small subset of areas and, as such, attention is increasingly turning to the automated analysis of the arguments in text: argument mining (Lawrence & Reed, 2020). Argument mining makes use of a variety of characteristic features of natural language that indicate argumentative intent. By combining such features and the accompanying techniques in a concerted approach, the insights from various disciplines and perspectives can be leveraged to achieve the best results (Lawrence & Reed, 2015).

In this chapter, we give a selective overview of the field of argument mining as the computer-based analysis of the logogic means of persuasion: the reasons advanced in justifying (or refuting) a disputed point of view.[2] After a short introduction to the methods of argument mining and argument corpora (Section 2), we discuss five aspects of argument mining in more depth: characteristic features of persuasive language (Section 3), rhetorical figures of speech (Section 4), types of statements used in persuasive text (Section 5), argument schemes as common forms of reasoning (Section 6), and persuasive dialogue (Section 7).

2 Methods of Argument Mining

The majority of argument mining techniques are based on some form of machine learning. Instead of requiring an *a priori* defined set of rules that the software applies to a given example of persuasive text, a machine learning approach makes use of a statistical model that learns from existing data. The system builds a model on the basis of an appropriately labelled set of examples (the *training data*), which is then tested on a yet unseen set of unlabelled examples (the *test data*) to determine how well the system is performing.

One of the challenges faced by current approaches to argument mining is the limited size and availability of appropriately labelled (or annotated) text corpora to serve as such training and test data. State-of-the-art techniques based on neural networks and deep learning especially require vast quantities of data to perform well and to prevent the system from over-fitting to an arbitrary

[1] http://www.stanford.edu/~rhorn/a/topic/phil/artclTchngPhilosphy.html

[2] While argument mining systems tend to be aimed at logos, there are some developments to also consider the automated mining the persuasive means of ethos, appeals to one's character, and pathos, appeals to the audience's emotions – see for instance Strommer (2011), Duthie and Budzynska (2018), and the present volume.

sample text, at the expense of wider applicability. Several recent efforts have been made to improve this situation through the creation of annotated text corpora and argumentation datasets across a range of different communicative domains. These efforts can be broken down into two main categories.

Manually annotated corpora consist of existing natural language texts with explicitly labelled argument components and structure. The Argument Annotated Essays corpus (AAEC2), for instance, comprises some 150,000 words annotated with premise-claim relations (Stab & Gurevych, 2017), while the US2016 corpus of presidential election debates (100,000 words) also includes annotation of dialogue structure and intertextual correspondence with social media posts (Visser, Duthie, et al., 2018).

An alternative to the annotation of existing texts is the generation of novel corpora from guided communication where the argumentative function can be inferred directly from the structure or script. The foremost example of this approach is the Microtext corpus of Peldszus (2014). This 8,000-word corpus of short persuasive essays was generated by tasking participants to write approximately five segments in which: all segments are argumentatively relevant; there is a segment acting as the main claim of the text; all other segments are supporting/attacking the main claim or another segment; and at least one possible objection to the claim is considered in the text. Whilst this method of generating textual data produces very clear examples of argumentation, the artificial nature of its construction means that results obtained on the dataset may not generalise well to naturally occurring unrestricted text.

3 Features of Persuasive Language

A strong indicator of argumentation in persuasive language is provided by discourse markers (van Eemeren et al., 2007), explicit linguistic expressions of the relationship between statements. For example, if we take the sentence "Britain should disarm because it would set a good example for other countries", then this can be split into two separate propositions "Britain should disarm" and "it [disarming] would set a good example for other countries". The presence of the discourse marker "because" connecting these two propositions clearly tells us that the second is being employed as a reason for the first. Discourse indicators (such as "because" and "however") have been successfully used as a component of argument mining techniques. For example, Stab and Gurevych (2014) used indicators as a feature in multiclass classification of argument components, with each clause classified as major claim, claim, premise or non-argumentative. There has, however, been little study of how well indicators perform on their own, how frequently they occur in real-world text, and how well different individual indicators map to specific argumentative relations.

There are many different ways in which indicators can appear, and a wide range of relations which they can suggest (Knott, 1996). Lawrence and Reed (2017) show how discourse markers can be used to automatically find argumentative sentences and indicate premise-conclusion pairs. They limit their search to specific terms indicating support or attack relations between pairs of propositions. Specifically, they consider those indicators which show an argumentative relation between sequential propositions of the form 'A–*indicator*–B' (e.g. "Britain should disarm *because* it would set a good example for other countries") or '*indicator*–A–B' (e.g. "*Because* we want to set a good example for other countries, we should reduce our nuclear capability"). They also consider the relationship between indicators and the directionality of the argumentative connections (e.g. 'A–*because*–B' suggests a support relation from the premise B to the conclusion A, whereas 'A–*therefore*–B' suggests a support relation from premise A to conclusion B). In this work, two sources of candidate discourse indicators were used: an aggregation of those found in existing literature (Groarke et al., 1997; Knott, 1996) (shown in Table 1), and a domain specific list extracted from the *US2016* corpora (Visser, Konat, et al., 2019). In both cases these lists were subsequently extended by including common synonyms.

Relation Type	**Indicators**
$A \xrightarrow{support} B$	so, therefore, accordingly, then, thus, consequently, hence, ergo
$A \xleftarrow{support} B$	because, since, as
$A \xrightarrow{conflict} B$	but, however, nonetheless, nevertheless, still, yet, though, whereas
$A \xleftarrow{conflict} B$	although, except, despite, albeit

Table 1: Argumentative discourse indicators from existing literature.

The results show that indicators which are commonly mentioned in the literature as being useful for identifying argumentative structure rarely occur in the examined data. Despite the possible influence of genre and activity type, it is surprising, for example, that the indicator "therefore" only occurs once within the entire US2016G1tv corpus[3]. This single instance, however, does indeed connect two inferentially related text spans. Of those indicators which appear more frequently in US2016G1tv, most provide little information. For example, whilst there were 30 instances of the indicator "so" occurring between adjacent spans, only 37.5% of these instances were between spans where a support relation exists. A possible explanation for this discrepancy can be found in the spoken genre of the US2016G1tv corpus, in which "so" may be used as a linguistic device signalling turn-taking instead.

The one exception here is the indicator "because". This indicator appears between text spans 71 times and, of these, 87.3% indeed indicated a support relationship. Whilst this is a promising result, suggesting that in those cases

[3]US2016G1tv is one of the *US2016* corpora (Visser, Konat, et al., 2019); specifically, analysis of the first general televised debate between Donald Trump and Hilary Clinton for the 2016 US presidential elections

where "because" occurs, it can tell us with high accuracy the type of connection, it is also shown that using this method on its own would leave approximately 80% of support relations unidentified (as well as all conflict relations).

These results are supported by those of earlier work which Lawrence and Reed (2015) carried out on the Araucaria corpus (Reed, 2006). Focusing on the thirteen most reliable support indicators and eleven most reliable conflict indicators, they discovered that in 89% of cases where one of these indicators occurred it corresponded to a support/attack relation, however only 4% of support/attack relations were explicitly marked in this way. Lawrence and Reed concluded that: "discourse indicators may provide a useful component in an argument mining approach, but, unless supplemented by other methods, are inadequate for identifying even a small percentage of the argumentative structure".

Fortunately, there are other features of persuasive language that can provide additional clues about the underlying argument structure. In particular, various measures of the similarity between propositions have been used to identify their argumentative relationship.

It seems intuitive that a premise and its associated conclusion may often have a large number of words in common (as in Example (1) from the US2016G1tv corpus), or be semantically similar without sharing very many common words (as in Example (2)).

(1) *Premise:* they lost plenty of money on investing in a solar company
Conclusion: that was a disaster to invest in a solar company

(2) *Premise:* We also have to make the economy fairer
Conclusion: I also want to see more companies do profit-sharing

Lawrence and Reed (2015) take advantage of this connection between the similarity of two text spans and their argumentative relationship. They use WordNet[4] to determine the semantic similarity between propositions and then consider the similarity scores for consecutive pairs. It is then assumed that if a proposition is similar to its predecessor then there exists some argumentative link between them,[5] whereas if there is low similarity between a proposition and its predecessor, the author may be going back to address a previously made point and, in this case, the proposition is compared to all those preceding it to determine whether they should be connected. This assumes that the argument is built up as a tree structure in a depth-first manner, where an individual point is pursued fully before returning to address the previous issues. Although the assumption of a tree structure does not hold for all arguments, it is the case for around 95% of the argument analyses contained in AIFdb.

A similar approach of assuming a relationship between argument components if they refer to the same concepts or entities is used by *AFAlpha* (Carstens et al., 2014), which represents customer reviews as trees of arguments, where a child-

[4]http://wordnet.princeton.edu/

[5]Further processing would be required to determine the direction of this connection (e.g. which is premise and which conclusion

parent relationship between two sentences is determined if they refer to the same concepts, with the child being the sentence that has been posted later. A sentence is represented as a set of features, including its semantic characteristics such as metadata about the review in which the sentence appears, as well as features based on the sentence's syntactic and lexical nature such as occurrences of certain words and phrase types. Each pair of sentences is then classified using a model trained on a data set comprised of data taken from the Q&A debating platform Quaestio-it[6], and the Internet Movie Database, IMDB[7].

The important role played by similarity is also exploited by Gemechu and Reed (2019), who borrow notions of aspect, target concept and opinion from opinion mining, and use these to decompose text spans down into finer-grained components, and then use similarity measures between these components to identify argument relations.

Finally, Wachsmuth et al. (2018) highlight an interesting link between similarity and argumentative relations. The work presented aims to determine the best counterargument to any argument without prior knowledge of the argument's topic. The best performing model rewards a high overall similarity between a potential counterargument and the given argument's conclusion and premises whilst punishing those counterarguments that are similar to only one of them. To some extent, this result captures the intuition that argumentative relations occur where something different is being said about the same topic.

4 Rhetorical Figures

The way an argument is presented can have a great impact on its persuasiveness. More than mere embellishments, linguistic devices such as figures of speech are often used to realise a specific rhetorical effect. Fahnestock (1999) makes a compelling case for the conception of rhetorical figures as couplings of linguistic form and communicative function. Drawing on the Aristotelian tradition, linking figures to *topoi*, Fahnestock argues that figures "map function onto form or perfectly epitomize certain patterns of thought or argument" (p. 26). This, in Fahnestock's terms, "figural logic" exhibits great promise for potential applications in argument mining and other computational approaches to argumentative discourse. If rhetorical figures are systematically connected to argumentative functions, then the presence of rhetorical figures in a text can be used as an indicator of argumentation, especially if the figure in question can be easily identified computationally.

Developing methods for the automated identification of rhetorical figures requires a compilation of the various types of figures with their definitions in a uniform notation. The most complete such compilation of rhetorical figures of which we are aware is the online resource *Silva Rhetoricae* (rhetoric.byu.edu) maintained by Burton (2017). The website assembles a large and diverse set of

[6] http://www.quaestio-it.com
[7] http://www.imdb.com

figures with descriptions, examples and detailed cross-referencing. An alternative, in some sense richer and more ambitious, but also as yet less comprehensive, online resource is the RhetFig website maintained at the University of Waterloo (artsresearch.uwaterloo.ca/rhetfig) (Kelly et al., 2010). Because of their digital format and comprehensiveness, both of these Web-based resources serve as invaluable starting points for the exploration of rhetorical figures in the context of what might be called 'computational rhetoric'.

The first work on computational rhetoric was at the *Symposium on Argument & Computation* held in 2000 in the Scottish Highlands (Reed & Norman, 2003). In the volume that was the output of the symposium, Crosswhite et al. (2003) explore how basic tenets of rhetoric might be accounted for in computational systems. They start with the classical distinction between *ethos*, *pathos* and *logos* as three means of persuasion, and then take aspects of Perelman and Olbrechts-Tyteca's (1969) *New Rhetoric* in particular, and formalise them using a context logic. Grasso (2002) develops these starting points further, with a particular view to building natural language generation systems that make explicit use of insights from rhetoric.

While both of these works include passing mentions of rhetorical figures, and Grasso emphasises their importance for "creating a computational model of rhetorical argumentation" (2002), neither gives any indication of how such a project might proceed. The first work to impose some computational order on figures was initiated by R. A. Harris and Di Marco (2009). This ongoing thread of research has more recently started to deliver results that can lay a foundation for detailed text processing (see, e.g., Mladenović and Mitrović's (2013) application to Serbian, and Gladkova's (2015) broader 'conspiracy of features' approach to argument diagnostics), and the mining of rhetorical figures (Gawryjolek et al., 2009; Hromada, 2011).

Crosswhite et al. (2003), Grasso (2002), and R. A. Harris and Di Marco (2009), amongst others, have demonstrated the great potential for the computational detection of rhetorical figures. It remains nevertheless an underdeveloped area of the computational approach to persuasive and argumentative discourse. Any foray, therefore, into the computational detection of figures could be a valuable contribution to argument mining, and argumentation studies and rhetoric in general (whether computational or not). In a special issue of the journal *Argument & Computation* on 'Rhetorical figures, argument, computation' (R. Harris & Di Marco, 2017), Lawrence et al. (2017) report on a pilot study looking at the relation between arguers' usage of rhetorical figures and the structure of their argumentation. Their starting point is that rhetorical figures as linguistic devices used to present moves in argumentative discourse can be expected to systematically relate to specific argumentation structures (following the form-function coupling of figures described in rhetorical scholarship).

While it remains difficult in argument mining to automatically recognise the structure of argumentation, it is relatively easy to identify the formal dimension of some rhetorical figures with automated techniques. This identification of figural forms can then be harnessed for argument mining as a feature in machine learning techniques, or equally serve in combination with other linguistic cues, domain knowledge, and common patterns of reasoning to mirror the complex process followed by a human annotator in reconstructing the structure of argumentation. Underlying the assumed connection between the textual structure characteristic of rhetorical figures, and the structure of argumentative is the claim in Fahnestock's figural logic that specific figural *forms* are systematically associated with corresponding figural *functions* (Fahnestock, 1999) (see also Tindale (1999) and R. A. Harris (2013)). In other words, a figure can be thought of as exhibiting a formal dimension and a functional dimension, with the systematic association of a particular form and a particular function constituting a particular type of rhetorical figure. This assumption forms the basis of the approach to improving the automated search for argumentation structures (related to figural *functions*) by means of searching for dialogical structures (instances of figural *forms*).

For example, we can consider *Epanaphora*, a figure of repetition describing the use of the same word or group of words at the beginning of successive clauses, sentences, or lines. Epanaphora is most commonly used to lend emphasis to these lines, making them more pleasurable to read and easier to remember as in Example (3) from the MM2012 corpus of BBC Radio 4 The Moral Maze transcripts.[8]

(3) It's not necessarily a moral good. *It can be* a way of demonstrating independence. *It can be* risk averse. *It can be* conservative.

Comparing such examples to the manual analyses of the argumentative structure shows that epanaphora can be a good indicator of several premises working together to support a conclusion – in this case, "It's not necessarily a moral good" is being supported by the three alternatives given. Indeed, for the 21 nodes identified as epanaphora in the MM2012 corpus of annotated *Moral Maze* episodes, 33.3% have a sibling relationship, significantly above the rate at which siblings occur in the corpus as a whole (15.1%).

Lawrence et al. (2017) apply the computational techniques for figure detection in a pilot study, arguing that the approach can be extended for a more detailed empirical testing of the specific correspondences between figural forms and functions as claimed in the rhetorical literature. Finding an instance matching the form of a particular rhetorical figure does not necessarily mean that the figure itself can be said to be activated: when the form occurs without the function (and similarly vice versa), this might only be called a quasi (or *prima facie*) instance of the rhetorical figure. The claims about specific form-function pairings of the rhetorical figures should however be grounded in empirical observation. Quantitative evidence could therefore suggest that some of the qualitatively demonstrated pairings should be revised, and the approach explored in the pilot study suggests a method for the production of such data.

The primary objective, however, was to explore the harnessing of figural forms as a resource for argument mining. In their pilot study, Lawrence et al. (2017) have demonstrated that the forms of rhetorical figures can be identified

[8] corpora.aifdb.org/mm2012c

automatically by employing the very definition of the figures as the algorithm for their identification. On the assumption that the figures indeed functionally contribute to persuasive argumentation, it should be possible to use the formal dimension of rhetorical figures as, for instance, a feature in machine learning approaches to argument mining.

5 Statement Types

Typologies of statements or propositions have traditionally played a central role in the study of persuasive language and argumentation. The three genres of rhetorical argumentation distinguished by Aristotle (1991) in his *Rhetoric*, for instance, can in part be characterised on the basis of the statement types predominant in each. Judicial rhetoric considers *factual* statements in a legal case about past events in determining guilt or innocence of an agent. Epideixis – or ceremonial oratory – tends to engender *evaluative* statements to praise or criticise behaviours and character traits. Lastly, deliberation in political discourse sees a predominant use of *policy* proposals appealing to advantages and disadvantages of future actions.

In the modern study of argumentation and persuasion, the types of propositions and their verification when used as premises in an argument are crucial issues (Freeman, 2000), leading to a variety of typologies of statements. For example, Reynolds and Reynolds (2002) distinguish between statistical, testimonial, anecdotal and analogical evidence; while Hoeken and Hustinx (2003) use a revised distinction between individual examples, statistical information, causal explanations, and expert opinions; Fahnestock and Secor (1988) employ the classical stasis issues of fact, definition, cause, value, or action; and Wagemans (2016) follows standards from debate theory by distinguishing values, policies and facts.

Such typologies have also been used in the annotation of text corpora for the purpose of training automated systems for classifying statements by means of machine learning. Park and Cardie (2014) classify propositions as unverifiable, verifiable non-experiential, or verifiable experiential – where verifiable propositions constitute assertions of objectively verifiable truth values. In a follow-up study, the authors (Park & Cardie, 2018) develop a corpus annotated with five proposition types: fact, testimony, value, policy and reference. Dusmanu et al. (2017) distinguish between factual and opinion-based arguments on Twitter, while Hua and Wang (2017) classify argumentative evidence as based on (statistical) studies, facts, opinions, or reasoning. Visser, Lawrence, Wagemans, et al. (2019) annotated the first television debate between Hilary Clinton and Donald Trump in the lead-up to the 2016 US presidential elections with statements of fact, policy, and value.

While the general classification of statement types is important in evaluating persuasive texts and, e.g., reconstructing argument schemes (see the next section), the classification of factual statements has gained particular prominence in recent years as part of fact-checking measures to combat fake news.

For developing manual and automated procedures for fact-checking alike, it is paramount to first know which statements can or should be checked. It only makes sense to *fact*-check *factual* statements, not, for instance, evaluative or policy statements.[9] To this end, Hassan et al. (2015) classify sentences as non-factual, unimportant factual, and check-worthy factual. Similarly, Patwari et al. (2017) and Jaradat et al. (2018) automatically determine the fact-check-worthiness of factual claims in political debates. Naderi and Hirst (2018), finally, automatically distinguish between true, false, stretch, and dodge statements in parliamentary proceedings. Given the role emotive language plays in the spread and success of fake news (Bakir & McStay, 2018), the automatic classification of value statements could also be leveraged in identifying and combating it.

6 Argument Schemes

Argument schemes capture persuasive inferences from a set of premises to a conclusion, relying on stereotypical patterns of (typically presumptive) human reasoning. As such, argument schemes represent a historical descendant of Aristotle's *topoi* (Aristotle, 1958) and, much like topoi, play an equally valuable role in the production, analysis, and evaluation of arguments – whether by human arguers or by automated software (see Chris Tindale's chapter 'Logos, Figuration and Schemes' in this volume). Several attempts have been made at creating taxonomies of the most common schemes (Garssen, 2001). Although these sets of schemes overlap in many places, the number of schemes identified and their granularity varies greatly. As a result, most argument analyses tend to contain examples from only one scheme set, with various permutations of the schemes described by Douglas Walton to be the most commonly used in computational approaches.

Understanding the argument scheme instantiated in a piece of natural language text can help us understand its persuasive force beyond what many existing automated techniques for extracting meaning offer. If we consider the product review shown in Example (4), then sentiment analysis techniques, for instance, allow us to understand at a high level what views are being presented – that this review is positive, for example – but they are unable to provide details on exactly why the reviewer recommends the product.

(4) The PowerShot SX510 is a fantastic camera. It's a Canon after all.

Analysing the argumentation in this review, we understand it to be enthymematic, with the implied missing premise "All Canon cameras are fantastic" surfacing in the reconstruction of the full structure of the argument. The two propositions "The PowerShot SX510 camera is made by Canon" (a reconstruction of the colloquial "It's a Canon after all") and "All Canon cameras are fantastic" are working together as a linked argument (Freeman, 2011;

[9] Which is not to imply that *fact-checking* alone is enough to combat fake news: we have argued elsewhere that *reason-checking* is another important component of the critical literacy involved (Visser, Lawrence, & Reed, 2020).

Snoeck Henkemans, 1992) to support the conclusion "The PowerShot SX510 is a fantastic camera". Furthermore, we can classify the inferential link between the premises and conclusion as an instance of the *Argument From Sign* scheme (Walton, 2006). Walton's approach to argument schemes assigns a particular label to each component part of a scheme instance. For the Argument From Sign in Example (4), the scheme components are as follows:

Specific Premise: The PowerShot SX510 camera is made by Canon
General Premise: All Canon cameras are fantastic
Conclusion: The PowerShot SX510 is a fantastic camera

The features of these common patterns of argument provide us with a way in which to both identify that an argument is advanced and determine its structure. By using the specific nature of each component proposition in a scheme, we can identify where a particular scheme is being used and classify the propositions accordingly, thereby gaining a deeper understanding of the argumentative structure which a piece of text contains.

Argument schemes can be a strong feature in argument mining and in the reconstruction of enthymemes (understood narrowly as arguments with unexpressed premises) (Feng & Hirst, 2011). To maximise their efficacy, the number and variation of individual schemes in annotated argument corpora should be as large as possible. Existing annotations, however, tend to use restricted sets of scheme types, while struggling to obtain reliable annotation results. For example, Duschl (2007) initially adopts a selection of nine argument schemes described by Walton (1996) for his annotation of transcribed middle-school student interviews about science fair projects. Later, however, he collapses several schemes into four more general classes no longer directly related to particular scheme types. This deviation from Walton's typology appears to be motivated by the need to improve annotation agreement. Similarly, Song et al. (2014) base their annotation on a modification of Walton's typology, settling on a restricted set of three more general schemes: *policy*, *causal*, and *sample*, while Anthony and Kim (2015) employ a bespoke set of nine coding labels modified from the categories used by Duschl (2007) and nine schemes described in a textbook by Walton (2006).

Visser, Lawrence, Reed, et al. (2020) develop an annotation procedure that aims to stay close to Walton's original typology while facilitating the reliable annotation of a broad range of argument schemes. The resulting method is reported to yield an inter-annotator agreement of 0.723 (in terms of Cohen's (1960) κ) on a 10.2% random sample of the annotated US2016G1tv corpus (Visser, Konat, et al., 2019). The main principle guiding the annotation is the clustering of argument schemes on the basis of intuitively clear features recognisable for annotators (Lawrence, Visser, Walton, et al., 2019). Due to the strong reliance on the distinctive properties of arguments that are characteristic for a particular scheme, the annotation procedure bears a striking resemblance to methods for biological taxonomy – the identification of organisms in the various sub-fields of biology. Drawing on the biological analogue, they propose a taxonomic key for the identification of argument schemes in accordance with Walton's typology: the Argument Scheme Key (ASK).

The ASK is a dichotomous identification key that leads the analyst through a series of disjunctive choices based on the distinctive features of a 'species' of argument scheme to the particular type. Starting from the distinction between source-based and other arguments, each further choice in the key leads to either a particular argument scheme or to a further distinction. The distinctive characteristics are numbered, listing between brackets the number of any not directly preceding previous characteristic that led to this particular point in the key. To classify the argument scheme in Example (4), for instance, the ASK leads the analyst through the following sequence: 1. Argument does not depend on a source's opinion or character; 17(1). Conclusion is not specifically action oriented; 32(17). Argument is not specifically value-based; 34(32). Argument is not about classification or legal rules; 45(34). Argument constitutes a single reasoning step; 47(45). Argument does not specifically rely on causality; 49(47). Argument relies on an individual case; 54(49). Argument refers to a characteristic sign; 55. Characteristic sign is present; 56. Premise mentions a sign: *Argument from Sign.*[10]

To further facilitate the annotation of argument schemes, Lawrence, Visser, and Reed (2019) develop a software solution that takes the user through the ASK in a series of binary choices to result in a suggested scheme. This ASK Assistant is integrated in the Online Visualisation of Argument (OVA) tool (Janier et al., 2014), a web-based application for analysing and annotating the argumentative structure of natural language text.[11]

7 Persuasive Dialogue

One of the most prominent applications of argument mining is found in computational systems for persuasive dialogue: AI systems designed to engage in spoken or written communication with one or more interlocutors, such as debates or other forms of argumentative interaction. These systems commonly use structured argument as a knowledge base from which the most persuasive points can be drawn at each speaking turn in the dialogue. While these systems have commonly used pre-annotated (Snaith et al., 2021) or crowd-sourced (Chalaguine & Hunter, 2020) data, attention is moving to systems backed by argument mining algorithms that can automatically harvest the most relevant data. The most ambitious of such systems is IBM's Project Debater (Slonim et al., 2021).

Following their previous Grand Challenges that resulted in DeepBlue and Watson, this latest bleeding-edge technology from IBM makes extensive use of

[10]The Argument Scheme Key can be thought of as a binary decision tree, which can be rendered textually (see Visser, Lawrence, Reed, et al. (2020)), or diagrammatically (see Lawrence, Visser, Walton, et al. (2019)).

[11]http://ova.arg.tech

Wikipedia and other data resources to create the first AI system that can debate humans on complex topics. Project Debater can respond to a given topic by automatically constructing a set of relevant pro/con arguments phrased in natural language. For example, when asked for responses to the topic "The sale of violent video games to minors should be banned", an early prototype of Project Debater scanned approximately 4 million Wikipedia articles and determined the ten most relevant articles, scanned all 3,000 sentences in those articles, detected sentences which contain candidate claims, assessed their pro and con polarity and then presented three relevant pro and con arguments[12]. Later versions of the AI system also work towards choosing the most convincing of these arguments (Gleize et al., 2019), expanding the topic of the debate (Bar-Haim et al., 2019), and providing "first principle" debate points, commonplace arguments which are relevant to many topics, where specific data is lacking (Bilu et al., 2019).

These argumentative abilities are the result of ongoing work in argument mining to extract meaningful argument data from large corpora. Levy et al. (2014) first addressed the challenge of detecting *Context Dependent Claims (CDCs)* in Wikipedia articles, showing how, given a topic and a selection of relevant articles, a selection of ("general, concise statements that directly support or contest the given topic") can be found. In following work, Rinott et al. (2015) addressed the extraction of supporting evidence from Wikipedia data for a given CDC. Bar-Haim et al. (2017) introduced the task of claim stance classification, that is, detecting the target of a given CDC, and determining the stance towards that target. Levy et al. (2017) further developed CDC identification, removing the need for pre-selected relevant articles, by first deriving a *claim sentence query* to retrieve CDCs from a large unlabelled corpus. Such large volumes of CDCs can be used both as potential points to be made by the Project Debater system as well as to aid in the interpretation of spoken material containing breaks, repetitions, or other irregularities (Lavee et al., 2019).

The method introduced by Levy et al. is used by Shnarch et al. (2018) to generate *weak labeled data* (data of low quality compared to manual annotation, but which can be automatically obtained in large quantities) and then combined with a smaller quantity of high quality, manually labeled data (*strong labeled data*). Using the combined strong and weak dataset resulted in improved performance for topic-dependent evidence detection, suggesting that this kind of data gathering can be a valuable asset, particularly in data-hungry neural network systems. The annotated datasets used in this and other Project Debater work are all available online.[13]

Another approach to argumentative dialogue systems is taken in the Centre for Argument Technology's *Arvina* and *Polemicist* applications. The web-based discussion software Arvina (Lawrence, Bex, & Reed, 2012) allows participants to debate a range of topics in real-time in a way that is structured but at the same time unobtrusive. Arvina uses dialogue protocols written using the

[12] http://www.kurzweilai.net/introducing-a-new-feature-of-ibms-watson-the-debater

[13] http://www.research.ibm.com/haifa/dept/vst/debating_data.shtml

Dialogue Game Description Language (DGDL) (Bex et al., 2014) to structure the discussion between participants. Such protocols determine which types of moves can be made (e.g. questioning, claiming, etc.), when these moves can be made (e.g. a dialogue starts with a claim; question moves can only made in the turn directly following a claim; etc.), and describe how each move updates the argument structure of the discussion taking place.

Arvina can support multiple human users interacting in the same dialogue, as well as incorporating software agents representing (the arguments of) specific authors who have their opinions stored in AIFdb (Lawrence, Bex, Reed, & Snaith, 2012). So, for example, say that Wilma has constructed a complex, multi-layered argument using the OVA argument analysis tool (Janier et al., 2014), concerning the use of nuclear weapons. A software agent representing Wilma can then be added to an Arvina discussion and questioned about these opinions, with the agent answering by giving Wilma's pre-annotated opinions.

The DGDL dialogue games available in Arvina have been developed to capture a variety of structured conversations, ranging from the collaborative generation of mathematical proofs (Pease et al., 2014) to debates on moral issues, and general discussion protocols to avoid fallacies (Visser, Budzynska, et al., 2018). The Polemicist[14] application a custom version of Arvina, giving users the opportunity to interact with software agents representing the participants of the BBC Moral Maze radio programme. Polemicist uses a fixed DGDL protocol, allowing the user to take on the role of the moderator of the debate: selecting topics, controlling the flow of the dialogue, and thus exploring all the angles of the rich argumentative content on display. Playing the role of moderator lets the user rearrange the arguments and create wholly novel virtual discussions between the contributions of participants that did not directly engage in the original debate, while staying true to their stated opinions.

Whilst Arvina and Polemicist currently rely on pre-annotated material from AIFdb to provide the responses of the software agents in a dialogue, they represent a valuable use case for automatically mined arguments. If the argumentative structure in a radio transcription can be extracted by argument mining, conversations in Polemicist could take place as the radio programme is being transmitted. Similarly, discussions aimed at creating new mathematical proofs in Arvina could feature counter-examples extracted from online mathematical publications. Combining these dialogue interfaces with a robust argument mining platform will enable users to discuss any issue of their choosing with any person whose opinions on that topic have been previously recorded.

8 Conclusion

The advent of argument mining has added new tools to the persuasion scholar's toolkit. Given the vast amounts of textual data available, the automated techniques provided by argument mining open up opportunities for quantitative research that are unavailable when relying on manual textual analysis, which is

[14]http://polemici.st

slow and prone to inconsistency or subjective interpretation. Extending existing opinion mining techniques, argument mining offers us the ability to rapidly process large amounts of textual data and determine not only *what* claims are advanced in persuasive communication, but also *why*, extracting the full argumentative structure at scale. Unsurprisingly, there is considerable commercial opportunity here too, as businesses increasingly want to build on the data that they gather in order to know more about the thoughts and behaviours of their customers, with many of the largest players in the field engaging – most visibly to date, IBM.

The availability of consistently annotated, high-quality argument data remains one of the main challenges to argument mining. Much recent work has focused on producing annotation guidelines targeted at specific domains, and whilst this has shown that data from these fields can be consistently annotated, the use of specific annotation schemes aimed at individual areas means that any techniques developed using this data are limited to the domain at hand. The volume of data, particularly data annotated at the most fine-grained level, is still far below what would be required to apply many of the bleeding-edge technologies brought to bear in other areas of computational linguistics. Attempts are being made to overcome this lack of data, including the use of crowd-sourced annotation (Ghosh et al., 2014) and automatic methods to extend the data currently annotated (Bilu et al., 2015). As these efforts combine with increasing attention to manual analysis, the volume of data available should increase rapidly.

Despite the relative sparsity of data when compared to other computational language tasks, argument mining techniques have been successfully deployed to extract details of the argumentative structure expressed within a piece of text, focusing on different levels of argumentative complexity as the domain and task require. Automated methods for analysing persuasive texts range from the basic identification of premises and conclusions using discourse markers and measures of semantic similarity, to, amongst others, the classification of statement types, argument schemes, and stylistic devices such as rhetorical figures, eventually providing the foundation for autonomous systems for persuasive dialogue.

Argument mining remains profoundly challenging, and traditional methods on their own seem to need to be complemented by stronger, knowledge-driven analysis and processing. However, the pieces required to successfully automate the process of turning unstructured textual data into structured argument analyses are starting to take shape. As the volume of analysed persuasive discourse continues to increase, and existing techniques are further developed and brought together, rapid progress can be expected.

References

Anthony, R., & Kim, M. (2015). Challenges and remedies for identifying and classifying argumentation schemes. *Argumentation*, *29*(1), 81–113.

Aristotle. (1958). *Topics* (W. D. Ross, Trans.). Oxford University Press.

Aristotle. (1991). *On rhetoric* (G. A. Kennedy, Trans.). Oxford University Press.

Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions. *Digital Journalism*, *6*(2), 154–175. https://doi.org/10.1080/21670811.2017.1345645

Bar-Haim, R., Bhattacharya, I., Dinuzzo, F., Saha, A., & Slonim, N. (2017). Stance classification of context-dependent claims. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, *1*, 251–261.

Bar-Haim, R., Krieger, D., Toledo-Ronen, O., Edelstein, L., Bilu, Y., Halfon, A., Katz, Y., Menczel, A., Aharonov, R., & Slonim, N. (2019). From surrogacy to adoption; from bitcoin to cryptocurrency: Debate topic expansion. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 977–990.

Bex, F., Lawrence, J., & Reed, C. (2014). Generalising argument dialogue with the dialogue game execution platform. In S. Parsons, N. Oren, C. Reed, & F. Cerutti (Eds.), *Proceedings of the fifth international conference on computational models of argument (comma 2014)* (pp. 141–152). IOS Press.

Bilu, Y., Gera, A., Hershcovich, D., Sznajder, B., Lahav, D., Moshkowich, G., Malet, A., Gavron, A., & Slonim, N. (2019). Argument invention from first principles. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1013–1026.

Bilu, Y., Hershcovich, D., & Slonim, N. (2015). Automatic claim negation: Why, how and when. *Proceedings of the 2nd Workshop on Argumentation Mining*, 84–93.

Burton, G. (2017). Silva rhetoricae: The forest of rhetoric. http://rhetoric.byu.edu

Carstens, L., Toni, F., & Evripidou, V. (2014). Argument mining and social debates. *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, 451–452.

Chalaguine, L., & Hunter, A. (2020). A persuasive chatbot using a crowdsourced argument graph and concerns. *Proceedings of the Eighth International Conference on Computational Models of Argument (COMMA 2020)*, 9–20.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Crosswhite, J., Fox, J., Reed, C., Scaltsas, T., & Stumpf, S. (2003). Computational models of rhetorical argument. In C. Reed & T. Norman (Eds.), *Argumentation machines: New frontiers in argument and computation* (pp. 175–209). Kluwer.

Duschl, R. A. (2007). Quality argumentation and epistemic criteria. *Argumentation in science education* (pp. 159–175). Springer.

Dusmanu, M., Cabrio, E., & Villata, S. (2017). Argument mining on twitter: Arguments, facts and sources. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2317–2322. https://doi.org/10.18653/v1/D17-1245

Duthie, R., & Budzynska, K. (2018). A deep modular rnn approach for ethos mining. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, 4041–4047. https://doi.org/10.24963/ijcai.2018/562

Fahnestock, J. (1999). *Rhetorical figures in science*. Oxford University Press.

Fahnestock, J., & Secor, M. (1988). The stases in scientific and literary argument. *Written Communication*, *5*(4), 427–443. https://doi.org/10.1177/0741088388005004002

Feng, V. W., & Hirst, G. (2011). Classifying arguments by scheme. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 987–996.

Freeman, J. B. (2000). What types of statements are there? *Argumentation*, *14*(2), 135–157. https://doi.org/10.1023/A:1007846431353

Freeman, J. B. (2011). *Argument structure: Representation and theory*. Springer.

Garssen, B. J. (2001). Argument schemes. In F. H. van Eemeren (Ed.), *Crucial concepts in argumentation theory* (pp. 81–99). Amsterdam University Press.

Gawryjolek, J., Di Marco, C., & Harris, R. A. (2009). An annotation tool for automatically detecting rhetorical figures system demonstration. *Proceedings of the IJCAI-09 workshop on Computational Models of Natural Argument*.

Gemechu, D., & Reed, C. (2019). Decompositional argument mining: A general purpose approach for argument graph construction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 516–526.

Ghosh, D., Muresan, S., Wacholder, N., Aakhus, M., & Mitsui, M. (2014). Analyzing argumentative discourse units in online interactions. *Proceedings of the First Workshop on Argumentation Mining*, 39–48.

Gladkova, O., Di Marco, C., & Harris, R. A. (2015). Argumentative meanings and their stylistic configurations in clinical research publications. *Argument and Computation*, *6*(3), 310–346.

Gleize, M., Shnarch, E., Choshen, L., Dankin, L., Moshkowich, G., Aharonov, R., & Slonim, N. (2019). Are you convinced? choosing the more convincing evidence with a Siamese network. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 967–976.

Grasso, F. (2002). Towards computational rhetoric. *Informal Logic*, *22*(3), 225–259.

Groarke, L., Tindale, C., & Fisher, L. (1997). *Good reasoning matters! : A constructive approach to critical thinking*. Oxford University Press.

Harris, R. A. (2013). Figural logic in mendel's experiments on plant hybrids. *Philosophy and Rhetoric*, *46*(4), 570–602.

Harris, R. A., & Di Marco, C. (2009). Constructing a rhetorical figuration ontology. *Symposium on Persuasive Technology and Digital Behaviour Intervention, Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB).*

Harris, R., & Di Marco, C. (2017). Rhetorical figures, arguments, computation. *Argument & Computation*, *8*(3), 211–231.

Hassan, N., Li, C., & Tremayne, M. (2015). Detecting check-worthy factual claims in presidential debates. *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, 1835–1838.

Hoeken, H., & Hustinx, L. (2003). The relative persuasiveness of different types of evidence. In F. H. van Eemeren, J. A. Blair, C. A. Willard, & A. F. Snoeck Henkemans (Eds.), *Proceedings of the fifth conference of the international society for the study of argumentation* (pp. 497–501). SicSat.

Hromada, D. H. (2011). Initial experiments with multilingual extraction of rhetoric figures by means of perl-compatible regular expressions. *Proceedings of the Second Student Research Workshop associated with RANLP*.

Hua, X., & Wang, L. (2017). Understanding and detecting supporting arguments of diverse types. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 203–208. https://doi.org/10.18653/v1/P17-2032

Janier, M., Lawrence, J., & Reed, C. (2014). OVA+: An argument analysis interface. In S. Parsons, N. Oren, C. Reed, & F. Cerutti (Eds.), *Proceedings of the fifth international conference on computational models of argument (comma 2014)* (pp. 463–464). IOS Press.

Jaradat, I., Gencheva, P., Barrón-Cedeño, A., Màrquez, L., & Nakov, P. (2018). Claimrank: Detecting check-worthy claims in arabic and english. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 26–30.

Kelly, A. R., Abbott, N. A., Harris, R. A., & Di Marco, C. (2010). Toward an ontology of rhetorical figures. *Proceedings of the 28th ACM International Conference on Design of Communication*, 123–130. https://doi.org/10.1145/1878450.1878471

Knott, A. (1996). *A data-driven methodology for motivating a set of coherence relations* (Doctoral dissertation). Department of Artificial Intelligence, University of Edinburgh.

Lavee, T., Orbach, M., Kotlerman, L., Kantor, Y., Gretz, S., Dankin, L., Jacovi, M., Bilu, Y., Aharonov, R., & Slonim, N. (2019). Towards effective rebuttal: Listening comprehension using corpus-wide claim mining. *Proceedings of the 6th Workshop on Argument Mining*, 58–66.

Lawrence, J., Bex, F., & Reed, C. (2012). Dialogues on the argument web: Mixed initiative argumentation with arvina. *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA 2012)*, 513–514.

Lawrence, J., Bex, F., Reed, C., & Snaith, M. (2012). AIFdb: Infrastructure for the argument web. *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, 515–516.

Lawrence, J., & Reed, C. (2015). Combining argument mining techniques. *Proceedings of the 2nd Workshop on Argumentation Mining*, 127–136.

Lawrence, J., & Reed, C. (2017). Mining argumentative structure from natural language text using automatically generated premise-conclusion topic models. *Proceedings of the 4th Workshop on Argument Mining*, 39–48.

Lawrence, J., & Reed, C. (2020). Argument mining: A survey. *Computational Linguistics*, *45*(4), 765–818. https://doi.org/10.1162/coli_a_00364

Lawrence, J., Visser, J., & Reed, C. (2017). Harnessing rhetorical figures for argument mining. *Argument & Computation*, *8*(3), 289–310.

Lawrence, J., Visser, J., & Reed, C. (2019). An online annotation assistant for argument schemes. *Proceedings of the 13th Linguistic Annotation Workshop*, 100–107.

Lawrence, J., Visser, J., Walton, D., & Reed, C. (2019). A decision tree for annotating argumentation scheme corpora. *3rd European Conference on Argumentation (ECA 2019)*, 97–114.

Levy, R., Bilu, Y., Hershcovich, D., Aharoni, E., & Slonim, N. (2014). Context dependent claim detection. *Proceedings of the 25th International Conference on Computational Linguistics*, 1489–1500.

Levy, R., Gretz, S., Sznajder, B., Hummel, S., Aharonov, R., & Slonim, N. (2017). Unsupervised corpus–wide claim detection. *Proceedings of the 4th Workshop on Argument Mining*, 79–84.

Metzinger, T. (1999). Teaching philosophy with argumentation maps review of can computers think? the debate by Robert E. Horn. *PSYCHE*, *5*.

Mladenović, M., & Mitrović, J. (2013). Ontology of rhetorical figures for Serbian. In I. Habernal & V. Matoušek (Eds.), *Text, speech, and dialogue: Proceedings of the 16th international conference, tsd 2013, pilsen, czech republic, september 1-5, 2013.* (pp. 386–393). Springer Berlin Heidelberg.

Naderi, N., & Hirst, G. (2018). Automated fact-checking of claims in argumentative parliamentary debates. *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, 60–65. http://aclweb.org/anthology/W18-5509

Palau, R. M., & Moens, M.-F. (2009). Argumentation mining: The detection, classification and structure of arguments in text. *Proceedings of the 12th international conference on AI and law*, 98–107.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135.

Park, J., & Cardie, C. (2014). Identifying appropriate support for propositions in online user comments. *Proceedings of the First Workshop on Argumentation Mining*, 29–38.

Park, J., & Cardie, C. (2018). A corpus of erulemaking user comments for measuring evaluability of arguments. In N. C. (chair), K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the eleventh international conference on language resources and evaluation (lrec 2018)*. European Language Resources Association (ELRA).

Patwari, A., Goldwasser, D., & Bagchi, S. (2017). Tathya: A multi-classifier system for detecting check-worthy statements in political debates. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2259–2262. https://doi.org/10.1145/3132847.3133150

Pease, A., Budzynska, K., Lawrence, J., & Reed, C. (2014). Lakatos games for mathematical argument. In S. Parsons, N. Oren, C. Reed, & F. Cerutti (Eds.), *Proceedings of the fifth international conference on computational models of argument (comma 2014)* (pp. 59–66). IOS Press.

Peldszus, A. (2014). Towards segment-based recognition of argumentation structure in short texts. *Proceedings of the First Workshop on Argumentation Mining*, 88–97.

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. University of Notre Dame Press.

Reed, C. (2006). Preliminary results from an argument corpus. In E. M. Bermúdez & L. R. Miyares (Eds.), *Linguistics in the twenty-first century* (pp. 185–196). Cambridge Scholars Press.

Reed, C., & Norman, T. (2003). *Argumentation machines: New frontiers in argument and computation* (Vol. 9). Kluwer.

Reynolds, R. A., & Reynolds, J. L. (2002). Evidence. In J. P. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 427–444). Sage.

Rinott, R., Dankin, L., Perez, C. A., Khapra, M. M., Aharoni, E., & Slonim, N. (2015). Show me your evidence-an automatic method for context dependent evidence detection. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 440–450.

Shnarch, E., Alzate, C., Dankin, L., Gleize, M., Hou, Y., Choshen, L., Aharonov, R., & Slonim, N. (2018). Will it blend? blending weak and strong labeled data in a neural network for argumentation mining. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 599–605.

Slonim, N., Bilu, Y., Alzate, C., Bar-Haim, R., Bogin, B., Bonin, F., Choshen, L., Cohen-Karlik, E., Dankin, L., Edelstein, L., et al. (2021). An autonomous debating system. *Nature*, *591*(7850), 379–384.

Snaith, M., Conway, N., Beinema, T., De Franco, D., Pease, A., Kantharaju, R., Janier, M., Huizing, G., Pelachaud, C., & op den Akker, H. (2021). A multimodal corpus of simulated consultations between a patient and multiple healthcare professionals. *Language resources and evaluation*, 1–16.

Snoeck Henkemans, A. F. (1992). *Analyzing complex argumentation*. SicSat.

Song, Y., Heilman, M., Beigman Klebanov, B., & Deane, P. (2014). Applying argumentation schemes for essay scoring. *Proceedings of the First Workshop on Argumentation Mining*, 69–78. https://doi.org/10.3115/v1/W14-2110

Stab, C., & Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 46–56.

Stab, C., & Gurevych, I. (2017). Parsing argumentation structures in persuasive essays. *Computational Linguistics*, *43*(3), 619–659.
Strommer, C. (2011). *Using rhetorical figures and shallow attributes as a metric of intent in text* (Doctoral dissertation). David Cheriton School of Computing University of Waterloo.
Tindale, C. (1999). *Acts of arguing: A rhetorical model of argument*. State University of New York Press.
van Eemeren, F. H., Garssen, B., Krabbe, E. C. W., Snoeck Henkemans, A. F., Verheij, B., & Wagemans, J. H. M. (2014). *Handbook of argumentation theory*. Springer.
van Eemeren, F. H., Houtlosser, P., & Snoeck Henkemans, A. F. (2007). *Argumentative indicators in discourse: A pragma-dialectical study*. Springer.
Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., & Reed, C. (2019). Argumentation in the 2016 US presidential elections: Annotated corpora of television debates and social media reactions. *Language Resources and Evaluation*, *54*, 123–154.
Visser, J., Lawrence, J., Reed, C., Wagemans, J., & Walton, D. (2020). Annotating argument schemes. *Argumentation*, *35*, 101–139. https://doi.org/10.1007/s10503-020-09519-x
Visser, J., Budzynska, K., & Reed, C. (2018). A critical discussion game for prohibiting fallacies. *Logic and Logical Philosophy*, *27*(4), 491–515. https://doi.org/10.12775/LLP.2017.021
Visser, J., Duthie, R., Lawrence, J., & Reed, C. (2018). Intertextual Correspondence for Integrating Corpora. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the eleventh international conference on language resources and evaluation (lrec 2018)* (pp. 3511–3517). European Language Resources Association (ELRA).
Visser, J., Lawrence, J., & Reed, C. (2020). Reason-checking fake news. *Communications of the ACM*, *63*(11), 38–40.
Visser, J., Lawrence, J., Wagemans, J., & Reed, C. (2019). An annotated corpus of argument schemes in US election debates. In B. Garssen, D. Godden, G. Mitchell, & J. Wagemans (Eds.), *Proceedings of the 9th conference of the international society for the study of argumentation (issa)* (pp. 1101–1111). Sic Sat.
Wachsmuth, H., Syed, S., & Stein, B. (2018). Retrieval of the best counterargument without prior topic knowledge. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, *1*, 241–251.
Wagemans, J. (2016). Constructing a periodic table of arguments. In P. Bondy & L. Benacquista (Eds.), *Argumentation, objectivity, and bias: Proceedings ossa 11* (pp. 1–12). OSSA.
Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Lawrence Erlbaum Associates.

Walton, D. (2006). *Fundamentals of critical argumentation.* Cambridge University Press.